R E P O R T R E S U M E S

ED 011 951 24 AA 000 070

CASE--A PROGRAM FOR SIMULATION OF CONCEPT LEARNING.

BY- BAKER, FRANK B.
WISCONSIN UNIV., MADISON

REPORT NUMBER BR-5-0216-11 PUB DATE 1 SEP 65
EDRS PRICE MF-$0.09 HC-$2.12 53P.

DESCRIPTORS- *CONCEPT FORMATION, *SIMULATORS, *LEARNING PROCESSES, COMPUTER PROGRAMS, PSYCHOEDUCATIONAL PROCESSES, INSTRUCTIONAL TECHNOLOGY, *INFORMATION SYSTEMS, *DATA PROCESSING, PROGRAMING, MADISON, CASE PROGRAM

THE "CASE" PROGRAM WAS DEVELOPED TO PROVIDE A VEHICLE FOR UNDERSTANDING THE PSYCHOLOGICAL PROCESSES INVOLVED IN CONCEPT LEARNING BY MEANS OF COMPUTER SIMULATION TECHNIQUES. BECAUSE THE MAJORITY OF PUBLISHED "SIMULATION OF CONCEPT LEARNING" PROGRAMS PROVIDED FEW INSIGHTS INTO THE LEARNING PROCESS, THE "CASE" PROGRAM WAS DESIGNED TO PROVIDE A BETTER MEANS FOR OBTAINING SUCH INSIGHTS. A PSEUDOCODE SCHEME WAS USED WITH A SPECIAL INTERPRETER WRITTEN IN MACHINE LANGUAGE WHICH PERMITTED SUBROUTINES TO BE USED IN SEVERAL CONTEXTS WITHOUT HAND CODING THE SITUATIONALLY DEPENDENT LINKAGES. A THREE-LEVEL MODEL OF HUMAN MEMORY INVOLVING WORKING MEMORY, SHORT-TERM MEMORY, AND LONG-TERM MEMORY WAS USED TO PROVIDE A FLEXIBLE MEANS FOR ACQUIRING, PROCESSING, AND STORING INFORMATION. "CASE" THUS REPRESENTS A SMALL PROGRAMING SYSTEM RATHER THAN A SPECIFIC COMPUTER PROGRAM AND, AS SUCH, CONTINUALLY CHANGES AS IMPROVED UNDERSTANDINGS ARE OBTAINED. AT THE TIME OF REPORTING, THE "CASE" COMPUTER PROGRAM WAS PRIMARILY A MEDIUM FOR EXPRESSING AND STORING THE INSIGHTS AND UNDERSTANDINGS OF THE CONCEPT LEARNING PROCESS WHICH HAVE BEEN ACQUIRED. THIS PAPER WAS PREPARED FOR PRESENTATION AT THE FALL JOINT COMPUTER CONFERENCE (LAS VEGAS, NOVEMBER 30 - DECEMBER 2, 1965). VARIOUS ASPECTS OF THE DEVELOPMENT OF THE PROGRAM ARE PRESENTED IN THREE PAPERS APPENDED TO THIS DESCRIPTION OF CASE. THESE PAPERS COVER (1) "CONCEPT ATTAINMENT EXPERIMENTATION BY COMPUTER SIMULATION," (2) " EXPERIMENTAL DESIGN CONSIDERATIONS ASSOCIATED WITH LARGE SCALE RESEARCH PROJECTS," AND (3) 'AN IPL-V TECHNIQUE FOR SIMULATION PROGRAMS." (JH)

BR-5.0216-10
P.A.24

AA000070

CASE: A Program for Simulation of Concept Learning.

Frank B. Baker

Research & Development Center for Learning & Re-Education

University of Wisconsin

Madison, Wisconsin

September 1, 1965

ED011951

Paper to be Presented at Fall Joint Computer Conference November 30-December 2, 1965, Las Vegas, Nevada.

CASE: A Program for Simulation of Concept Learning

Frank B. Baker
University of Wisconsin

The CASE program was developed to provide a vehicle for the understanding of the psychological processes involved in concept learning by means of computer simulation techniques. Because the majority of published simulation of concept learning programs provided few insights into the learning process, the CASE program was designed to provide a better means for obtaining such insights. A pseudo-code scheme is used with a special interpreter written in IPL-V which permits sub-routines to be used in several contexts without hand coding the situationally dependent linkages. A three-level model of human memory involving working memory, short term memory, and long term memory was used to provide a flexible means for acquiring, processing, and storing information. Because of the above, CASE represents a small programming system rather than a specific computer program and as such continually changes as improved understandings are obtained.

At the current time, the CASE computer program is primarily a medium for expressing and storing the insights and understandings of the concept learning process which have been acquired.

CASE: A Program for Simulation of Concept Learning

The Learning Research and Development Center of the University of Wisconsin is engaged in a long-term multi-facet study of concept learning, supported by the U. S. Office of Education. The concept attainment simulation experiment (CASE) is the facet of this overall effort which utilizes the technology of computer simulation as a vehicle for obtaining a better understanding of the psychological processes involved in the learning of concepts. The long-range goal is the utilization of the insights thus obtained to improve classroom learning. The study of concept learning has a long history within psychology and has received considerable attention in recent years due in part to the book by Bruner, Goodnow, and Austin (1956) which delineated strategies for learning concepts. The experimental materials used by Bruner consisted of a finite universe of objects each of which possessed n dimensions and each dimension could assume k different values. A classification rule (a concept) consisting of a particular combination of dimension values partitioned the universe into two mutually exclusive sets. In a typical experiment a subject was shown an object which was an exemplar of the set defined by the concept and told his task was to ascertain the classification rule. In order to attain the concept the subject chose objects from the universe and the experimenter indicated the set membership of the object chosen. The object selection-designation procedure continued until the subject could verbalize the correct classification rule and hence the concept had been attained. The experimental situation, the problem to be

solved, and learning procedure involved appear reasonably simple and a number of persons have written programs to simulate this type learning experiment, Hunt and Hovland (1960), Hovland and Hunt (1960), Hunt (1962), Allen (1962), Wickelgren (1962), and Baker (1964). The book by Hunt (1962) provides an excellent review of much of the psychological literature relevant to concept learning as well as discussing his own simulation program. Unfortunately the existing programs leave one with the disquieting feeling that although they attain concepts, little has been added to our understanding of the psychological processes involved in concept learning. Most of these programs are at best watered-down algorithms and involve very little of psychological importance. Because of the shortcomings of the existing simulation programs a project was initiated to develop a program which hopefully will eventuate in something of psychological significance.

The basic approach was to use Bruner's notions about learning strategies, coupled with concepts regarding the structure of behavior from the book by Miller, Galanter, and Pribam (1960) to write a computer program which would attain concepts. This initial program based on semi-theoretical grounds would then served as a stepping-off point for a learning process on the part of the present author.

A system for collecting data was established which consisted of a closed feedback loop, with the simulation program at one end and protocol gathering during experiments involving human learning at the other end. Within the computer program, certain routines may be based upon a priori grounds or represent areas not clearly understood.

In order to get better insights into such areas, questions are used during the protocol gathering which will elicit verbalizations relevant to those points. Thus, the computer program guides the production of information within the protocol which is subsequently used to modify the program itself. By making an extremely close connection between the computer program development and the learning experiments with human subjects the hope is to obtain a better understanding of the psychological processes involved. Having set the broad context within which the project operates, let us next turn our attention to the actual computer program involved.

The CASE Program

Memory Structure

During the early phases in the evolution of the CASE program it became obvious that one of the keys to the problem was an adequate representation of the structure of human memory. The psychological literature contains a considerable body of material related to memory and much of this was studied to ascertain an appropriate structural form of memory. The result of this search was to design a memory consisting of three levels: Working memory (WM), short-term memory (STM), and long-term memory (LTM). The working memory is a unit which serves two functions. One, it holds all information received from the external environment until it can be analyzed and re-coded for transmission to a more permanent level of memory. Second, it serves as a buffer memory for holding information which is created within the subject and must be passed from one information-processing routine to another. In this buffer mode it provides certain higher-level routines

contextual information which is used to guide program flow. The short-term memory is semi-permanent and retains information relevant to the current state in the learning of a particular concept. Short-term memory can receive inputs only from routines which re-code and transmit the contents of working memory or long-term memory. Long-term memory will contain information re-coded from short-term memory concerning concepts learned and how they were learned but at the present time only working memory and short-term memory have been programmed. Figure 1 illustrates the communication paths within the memory structure. The only means of communication from STM and LTM to the external world is via the output channel. For example, the subject tells the experimenter which object he has selected via this channel at the experimenters designation of the set membership of the object is received by the subject via working memory.

Figure 1 about here

The internal structure of short-term memory consists of lists having a somewhat unusual IPL-V structure which has proven extremely useful. The structure employs two levels of attributes; the class attributes which represent a rather broad description such as the permanent characteristics of an object; and specific attributes such as an objects serial position in the external environment, thus providing a detailed description within the class attribute. Table 1 illustrates a typical list within STM. The description list 9-0 describes M13, description list 9-1 describes the symbol E0 on M13. The description list 9-1 contains class attributes, such as A2, and its attribute value list V2 which is merely a storage device for symbols

whose function is to hold a description list containing the specific attributes and their values (A4, V4: A5, V5). The value list of the class attribute is a push-down list whose top symbol always represents current information.

Table 1 about here

Notice in Table 1 that the list structure is symetrical in form to the upper left and lower right of the dotted line. The symmetry enables one to write simple routines which function for a module of memory regardless of the level at which the module occurs within the structure. There are four such basic memory routines which do all of the STM input and output:

1. Remember a name
2. Remember something about that which has been named
3. Recall a name
4. Recall something about that which has been named

At the present time we have not attempted to include LTM or to introduce forgetting or interference, however, we anticipate at some point building such mechanisms into the memory structure.

Program Structure

The CASE computer program has been designed with an expandable hierarchical structure whose depth depends upon the level of sophistication obtained in understanding the learning process. At the present time there are four levels with each level being tested within the next higher level as an IPL-V list structure. The list structure representing the learning process is presented as input to a special interpreter (Baker and Martin, 1965) which executes the

symbols in the structure and performs a number of housekeeping functions. The upper level (S) specifies what Bruner, et al. (1956) refer to as a strategy, and is a list of symbols which represent major procedures within a strategy. The next lower level is the procedure level (Z-D) which is a list of symbols representing the processes combined to accomplish a given procedure, such as searching the external environment for an object having certain characteristics. The next lower level (P-Q) consists of information processing routines written in IPL-V and is the lowest level that the program can manipulate at run time. The fourth level (R) consists of basic information processing modules coded in IPL-V which a programmer can use to manually write new P-Q level routines. The R's are subroutines which do such things as compare, test for the presence or absence of information, etc. Within each level it is necessary to maintain a sharp distinction between routines which perform operations (Z's and P's) and those which provide decision-making information (D's and Q's). It has also been found necessary to defer decision making upward to the next higher level for action. It should be noted, that all levels above the P-Q and R levels merely consist of symbols, hence one writes IPL-V only at the lowest level - a fact which has many implications for how one studies the learning process.

In that the P-Q level routines are basic information-processing routines they can be used in a wide variety of situations within the program where the processing is identical but the information, its source, and its disposition differ. Because of this characteristic of the P-Q level routines one is faced with the problem of how to use

the same routine in a number of different contexts without developing a significant amount of situationally dependent IPL-V coding. The solution devised uses a pseudo code whose description list contains the inputs, outputs, and characteristics of the routine. The inputs to a routine are determined by a higher-level routine called a contexter which in the case of the P-Q level routines uses the contents of working memory to determine what the inputs to the processing routine should be for a given situation. Currently the outputs are normally placed in working memory although other options are possible. Although we have not done so yet it appears feasible to put on the description list of the routine a specification of the kind of information processing the routine is capable of performing. Table 2 shows a P level routine before and after the context program has functioned. Once the inputs have been determined from the context, the routine is returned to the interpreter for execution. Although it has not been done, the concept of context routines which operate at all levels within the program structure appears feasible and the context routine at the highest level would be a plan to create plans such as suggested by Miller, Galanter and Pribam (1960).

Table 2 about here

CASE Mark 3 Mod 1

The preceding discussion indicates the development of what might be loosely described as a programming system rather than a unique computer program such as LT, EPAM, or the previous concept attainment programs. The inherent flexibility in the system makes it difficult to discuss the CASE program per se. Because of this difficulty,

periodically a particular configuration is set aside, given a mark and mod number, and documented. At the time this paper was written CASE Mark 3 Mod 1 was the operational program which implemented what Bruner, et al. (1956) called the conservative focusing strategy. The implementation of this strategy required 9 routines at the Z-D level and 13 at the P-Q level. Table 3 lists some representative routines at these two program levels.

Table 3 about here

Fundamentally the program is still an algorithm as it very efficiently learns every concept attempted in a minimum number of object choices. Although it currently shares this fault with its published predecessors we feel its internal structure is more psychologically oriented and the potential for non-algorithmic behavior exists.

Some Retrospects

Program Characteristics

When one reviews the history of the CASE program it becomes quite clear that a subtle process is in effect. Namely, as one's understanding of the learning process increases the computer simulation program changes from routines which perform a large block of the concept attainment process to a number of short routines which can be widely employed. In the CASE program such a change has been dramatic at the P-Q level from Mark 1 Mod 0 to Mark 3 Mod 1. The cynic will counter that we are merely learning how to code IPL-V but I do not believe this is the only basis, as the change has been effected primarily on psychological grounds rather than coding considerations. In fact, separate symbols are used to designate routines which are

the result of IPL-V rather than psychological considerations and the former routines are quite rare. The character of the subroutines in the CASE program have also changed from being highly specific to the Bruner type experimental situation to being reasonably independent of the experimental situation. They are, however, dependent upon the basic memory structures defined earlier. The situationally dependent tasks still get performed but the computer program is problem specific at a higher level than was previously true.

Outcomes of the CASE Program

There have been a number of outcomes of the CASE effort which are as follows:

1. The hierarchical structure of the processing routines and what appears to be a parallel structure in context routines has led us into a continual search for logical units within the learning process. Behaviors which once seemed quite dissimilar have been decomposed and found to share a number of basic information processing modules. As a result we are slowly acquiring a better understanding of the psychological processes involved in concept learning.

2. The development of the CASE program has generated ideas for classical psychological experiments in a number of areas as a result of problems arising during the development of certain subroutines. Topics such as the role of dominant dimensions in a subject's learning behavior and the lack of independence among dimensions in the Bruner experimental materials have been elicited. It appears as though an important outcome of computer simulation is the generation of ideas which can be researched in the usual psychological experimental setting.

3. A completely unexpected outcome has been that we rarely make a production run on the computer, a fact which seems anomalous in a computer simulation project. What has happened is that enormous numbers of man hours have been devoted to gathering and studying protocols, to the development of programming techniques in order to implement the next level of sophistication within the system, and to analysis of the computer program itself. These activities plus the lack of variability in the learning behavior of the CASE program at this point in time have resulted in a relatively few production runs.

Conclusions

The CASE project has <u>not</u> been conceived as an effort which will produce immediate spectacular results, rather we view it as a slow developmental process. The memory structure, the program structure, the interpreter, and the contexter are basic concepts which we feel will enable us to continuously improve the sophistication of the program as our understanding of the concept learning process improves. Although it is difficult to single out specific accomplishments of great psychological importance, a certain modicum of progress has been made along these lines. At the current time the computer program is primarily a medium for expressing and storing the insights and understandings of the concept learning process which we have acquired. Uhr (1963) has previously indicated that psychological theories might be expressed in the form of computer programs and our experience to date tends to substantiate this point of view.

References

Allen, M., "A Concept Attainment Program That Simulates a Simultaneous-Learning Strategy," Behavioral Science, vol. 7, pp. 247-250 (1962).

Baker, F. B., "An IPL-V Program for Concept Attainment," Educational and Psychological Measurement, vol. 24, pp. 119-127 (1964).

Baker, F. B. and Martin, T. J., "An IPL-V Technique for Simulation Programs," Educational and Psychological Measurement, (In Press).

Bruner, J. S., Goodnow, J. J. and Austin, G. A., A Study of Thinking, John Wiley and Sons, Inc., New York, 1956.

Hovland, C. I. and Hunt, E. B., "Computer Simulation of Concept Attainment," Behavioral Science, vol. 5, pp. 265-267 (1960).

Hunt, E. B. and Hovland, C. I., "Programming a Model of Human Concept Formation," Proceedings of the Western Joint Computer Conference, pp. 145-155 (1960).

Hunt, E. B., Concept Learning: An Information Processing Problem, John Wiley and Sons, Inc., New York, 1962.

Miller, G. A., Galanter, E. and Pribam, K. H., Plans and the Structure of Behavior, Holt, Rinehart, and Winston, New York, 1960.

Uhr, L., "Pattern Recognition Computers as Models for Form Perception," Psychological Bulletin, vol. 60, pp. 40-73 (1963).

Wickelgren, W. A., "A Simulation Program for Concept Attainment by Conservative Focusing," Behavioral Science, vol. 7, pp. 245-247 (1962).

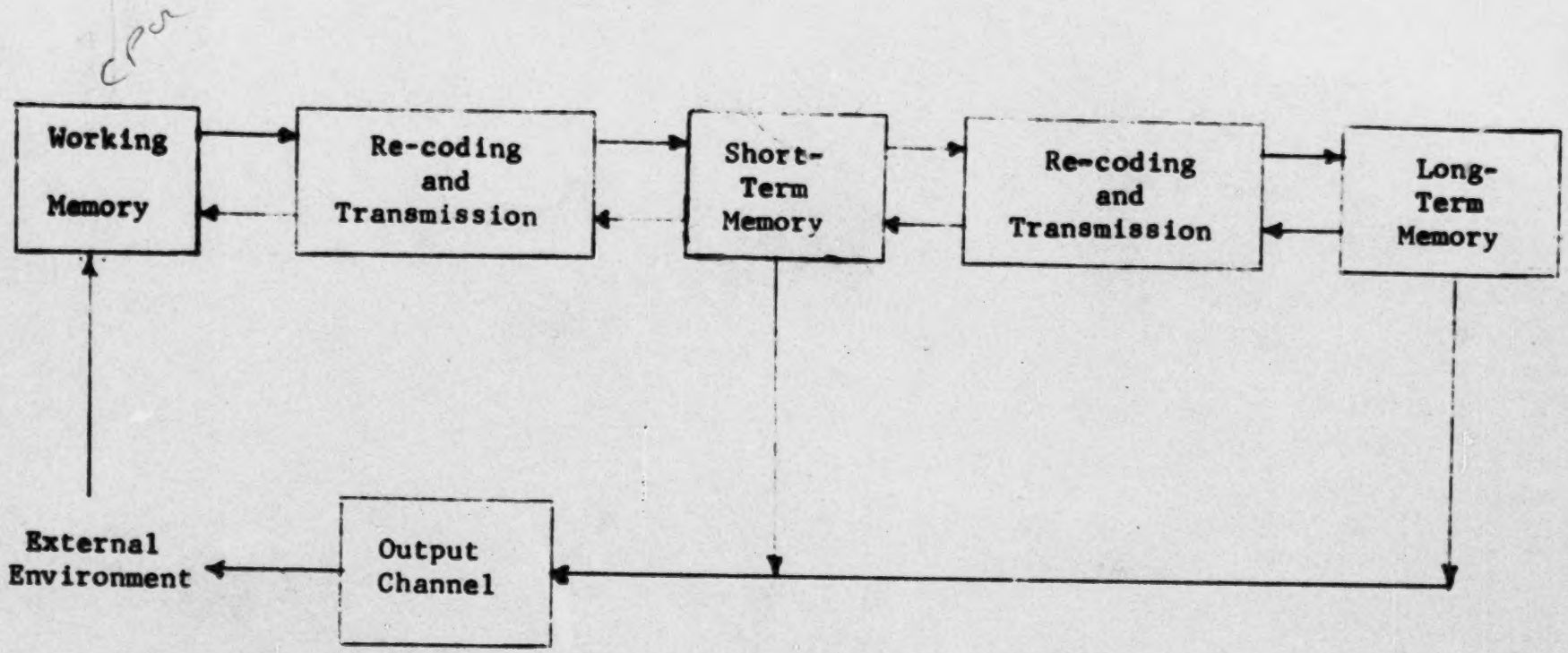

Figure 1. Communication Paths within CASE Memory Structure.

Table 1. Typical Memory List in Short-Term Memory

M10	9-0				
	E0				
	E1				
	E2	0			
9-0	0		E0	9-1	
	A1			X1	
	V1	0		X2	0
			9-1	0	
				A2	
				V2	
V2	9-2				
	Y2				
	Y1	0			
9-2	0		Y2	9-3	0
	A3				
	V3	0	9-3	0	
				A4	
				V4	
				A5	
				V5	0

Table 2. Pseudo Code System

Before Context Program				After Context Program			
P31	9-0		Pseudo Routine	P31	9-0		
	P30	0	Executable Routine		P30	0	
9-0	0			9-0	0		
	A7		Inputs		A7		
	V7	0	0		V7	M11	
	A8		Outputs			M12	
	V7	M1	0			A6	0
					A8		
					V7	M1	0

Table 3. Representative Routines Currently Implemented

Procedure Level
Z1 Form criteria for choosing an object
Z2 Find object in external environment meeting criteria
D1 Determine if concept should be offered
Z4 React to class membership of object
Process Level
P3 Vary value of a dimension
P6 Remember something about object named
Q1 Test for specific value of a given item of information
P12 Collect items having common attribute

CONCEPT ATTAINMENT EXPERIMENTATION BY COMPUTER SIMULATION*

Frank B. Baker
Research and Development Center
for Learning and Re-Education
University of Wisconsin

In order to set the context of the research effort described in the present paper, I would like to mention briefly the learning research and development center at the University of Wisconsin under whose auspices this work was performed. The Center is one of four recently funded by the U.S. Office of Education and has as its central theme the study of classroom learning with special emphasis upon concept learning. The activities of the Center include basic research, development and dissemination, the underlying rationale being that ideas would be taken from basic research and translated into classroom practices. Because of the dearth of basic research in the area of the learning process a few of us are engaged in what hopefully is some basic research employing the technology of computer simulation. Computer simulation of human behavior stems from the pioneering work of the group at Carnegie Tech and the RAND corporation since the mid-1950's. At the current time a wide variety of simulation research has been reported and the field is growing rapidly.

The study of concept learning by means of computer simulation techniques seemed to us to be a reasonable endeavor due to the rather straightforward experimental situations involved. We have found, however, that its simulation is not straightforward. Fundamentally, one cannot write a computer program for something which is ill defined, hence a major effort has been devoted to trying to model the concept learning process. What one attempts to do is to find a framework for the learning process and then fill in as many details as possible. The explicit nature of the computer and the vagueness of current knowledge makes this modeling process an extremely difficult task. The technique which we have

---

*Paper presented at Annual Convention of the American Educational Research Association, Chicago, February, 1965. The research reported herein was performed pursuant to a contract with the U.S. Office of Education, Department of Health, Education, and Welfare.

devised to deal with this modeling task is a closed loop feedback system consisting of the computer program at one end of the loop and laboratory experiments involving students at the other. The usual procedure employed by Simon and others in simulation research has been to use the "thinking aloud" procedure in which the subject describes what he is doing as the experiment progresses, a technique which is frowned upon by a large segment of the psychological community as smacking of introspection; however, very valuable data can be obtained in this manner. From a modeling point of view it has another disadvantage in that one obtains only what the subject happens to talk about. In our feedback loop we have employed a somewhat different strategy which uses the computer program as a guide. For example, in the typical Bruner type learning experiment a card on the board is designated as the focus card at the beginning of the experiment and presumably is remembered by the subject. But let us assume that at a certain point within the experiment an inaccurate memory of the focus card results in failure to attain the concept. The "thinking aloud" procedure enables the experimenter to follow the subject and when the critical juncture known from the computer program is reached, he interjects questions to ascertain if the subject still retains the focus card information. I might add that this probing might aid a subject to attain the concept which he otherwise might not attain, but that is immaterial to our interests. The heart of this procedure is that the questions asked are relevant to some aspect of the simulation program that we are trying to develop and the "thinking aloud" procedure provides us with the vehicle for interjecting the proper questions at the appropriate time. In addition, one does obtain the usual protocol information. When studying these protocols we try to ascertain for example if a particular subroutine exists within the protocol. Thus, the computer program helps guide our thinking when interpreting the verbal

behavior. The information obtained is then used to modify the computer program in such a way that it will correspond more closely with the experiments. What I have described above is the framework of an investigational procedure which I think will be of material aid in the modeling process, due to the fact it can yield highly specific information in a form usable in computer program development. We have made several runs with this procedure and I'd like to mention some initial impressions.

We have employed university sophomores as professional subjects who are used about one hour a week in the laboratory experiments, the rationale being that we will get some consistency of behavior by using the same people repeatedly and they will become adept at using the "thinking aloud" procedure. The questions are varied from problem to problem to counter-balance any learning effects relative to the type of questions asked. Our initial computer program was based on Bruner's wholist strategy and in order for it to attain concepts under this scheme it must remember certain types of information. The first problem which arose was to make a decision as to how people actually represent the experimental environment. In that the board consisted of only 32 cards I felt that the subjects would remember objects or at least where they were on the board; my graduate assistants disagreed and so off we went to run our ten subjects. A series of questions was devised to ascertain if the subjects dealt with cards as objects and if they maintained any form of a cognitive map as to their location. Much to my horror and the assistants' glee, some rather intensive probing elicited that the subjects did not treat the cards on the board as objects but rather they dealt in attributes and their values. At this juncture a major aspect of my simulation program went down the drain and we had only asked one question!! Back to the drawing boards!

It became readily apparent that this simulation business involves a lot of false starts and rather sudden changes in perspective which can readily obsolete

vast amounts of computer programing. Thus, a technological problem arose as to how to minimize the effect of major changes in the program and maximize the ability to make the changes. These two goals are not easily reconcilable; however, Mr. Tom Martin and myself feel we have a reasonable solution. We established an interpretive programing system within IPL-V (which is somewhat like bringing crime to Chicago) which recursively executes lists of program names and automatically takes care of the subroutine interface problem. In somewhat different terms, what we did was to establish an executive program which does not care what its subroutines happen to be. The basic computer program can operate without depending upon the information processing actually performed by the subroutines. Thus, the interpretive system serves as a super program above the simulation program. Within the actual simulation program there are four levels of programs which are themselves lists. The S or strategy level essentially is an executive level description of a particular learning strategy, the next lower level is the Z-D level which is the major procedure level, the third level down is the P-Q level which is the process level, and the R level is the module level which is the lowest unit within the program. The R level routines are the basic information processing capabilities which we provide our "subject", such as being able to compare, being able to locate, remembering, etc. The next level up consists of routines which do things (P's) and routines which make decisions (Q's), much in the same manner as parts of Miller's TOTE units except that we use only test and operate. The P's and Q's consist of a number of R's and some IPL-V machine language instructions, hence are the lowest level which gets disturbed by changes in the simulation program. The Z-D level consists of the major blocks in the learning strategy and performs such things as generating the basis for an object choice or deciding what to do after an object has been designated by the experimenter. Note that the distinction between doing and deciding is maintained at all levels and that decisions are always deferred

up to a higher level for action. The Q's result in action at the Z level and the D's cause action at the S level. Such an upward communication is necessary if one is to maintain proper control of the decision-making process.

Assuming that sufficient P's and Q's are available, the writing of a simulation program consists merely of connecting P's and Q's to form Z's and D's, then connecting Z's and D's to form an S list. The interpreter then proceeds back down the list structure to execute each symbol in turn and hence perform the simulation. Table 1 presents the wholist strategy in the scheme.

It is interesting to note that the names of the subroutines almost form a verbal description of the concept attainment strategy, a fact which offers some interesting possibilities for a string language notation. We have explored the recursive execution of such lists and the programing technique seems to be correct. Two related problems were also resolved while developing the interpreter. First, if one can remove, insert, or rearrange subroutines at will within such a scheme, a difficult interface problem arises from the need for the interpreter to know what inputs/outputs will be created. Mr. Martin and I resolved this by making each subroutine carry on its own description list a specification of the inputs and outputs for that routine. Thus, before each routine is executed the interpreter can ascertain what the inputs are, obtain them and set up the entering arguments and in one fell swoop the interface problem disappeared. Currently, the human programmer must make sure the inputs exist; however, someday the interpreter will search to see if the necessary inputs have been created and stored. I might add I have to constrain my programmer from trying to write a program which selects the appropriate program for insertion, deletion, etc. At the present time, we have not been able to fully explore the possibilities of the interpreter, as our laboratory experiments have been running painfully slow and the generation of new P's, Q's, and R's has not kept pace with the computer technology; however, the interpretive program is debugged and we foresee no real programing problems.

Table 1: Typical Concept Simulation Program List

```
Wholist Strategy List
     S1  9 - 0  Program Description List
         X1     Initialize Problem
     9-1 F1     Form Hypothesis
     9-2 F2     Select Object
         D1     Object Usable?
         9 - 2  No
         F3     Yes, designate by Experimenter
         D2     Proceed?
         9 - 3  No
         F4     Present Concept
         D3     Concept Correct
         9 - 1  No
         X2     Print History of Problem Solution
     9-3 0      End

Typical Second Level Program List
     F1  9 - 0  Program Description List
     9-1 P1     Create a n valued Hypothesis
         Q1     Previously Used?
         9 - 3  No
     9-2 Q2     Change Dimensionality?
         9 - 1  No
         P2     Yes, change Dimensionality
         9 - 1
     9-3 Q3     Intersection OK
         9 - 2  No
         0      Yes, End
```

A review of our effort over the past year indicates that our current approach to programing via the interpreter scheme seems to offer a high degree of flexibility. We have attempted to design a basic system which will not change as the problem is expanded, and our present feeling is that we have attained this goal. The really serious problems have occurred in the protocol-gathering area, as we have not been able to devise a scheme for rapid transcription of the experimental situation. We can generate new avenues of exploration and conduct the experiments much faster than we can document the sessions. In that our ultimate goal is trying to understand the learning process we are striving to get around these technological problems as quickly as possible and get to the main problem. But as you can see, it's not easy.

References

Baker, F. B. "An IPL-V Concept Attainment Program." Educational and Psychological Measurement. Spring, 1964, 24, 119-127.

Bruner, J. S., Goodnow, J. J., and Austin, G. A. A Study of Thinking. New York: John Wiley and Sons, Inc., 1956.

Hunt, E. B. Concept Learning: An Information Processing Problem. New York: John Wiley and Sons, Inc., 1962.

Miller, G. A., Galanter, E. and Pribram, K. H. Plans and the Structure of Behavior. New York: Holt, Rinehart, and Winston, 1960.

Newell, A. (Ed.) Information Processing Language-V Manual, Second Edition. Englewood Cliffs, New Jersey: Prentice-Hall Inc., 1964.

Simon, H. A. and Kotovsky, K. Human Acquisition of Concepts for Sequential Patterns. Psychol. Rev., 1963, 70, 534-546.

Experimental Design Considerations Associated with Large Scale Research Projects*

Frank B. Baker

Laboratory of Experimental Design
and
Center for Research and Development in Learning and Re-education

University of Wisconsin

The majority of papers presented at this symposium have dealt with technical aspects of experimental design, but the current paper does not aspire to this end. Rather, the intent is to examine experimental design in relation to the type of research programs currently being undertaken. For quite some time, the author has had a suspicion that the characteristics of problems presently being studied are quite dynamic in nature, whereas the conceptualization of experimental design employed in educational research has become rather static. In order to pursue this premise, let us first examine the basis for experimental design as we now understand it.

During the 1920's and 1930's, agricultural experimentation was approaching its zenith and the agricultural experiment stations which existed throughout the United States and Europe were the foci of considerable research. The agricultural research worker of this era was interested in developing crops having better yield, breeding animals which gained more weight, applying fertilizers in a manner which would increase yield, and other such projects. These research projects had a number of characteristics which made them particularly amenable to statistical analysis. They were essentially univariate in that a variable such as weight, bushels per acre, size of ham, etc., was readily available to provide a quantitative measure which could be used for mathematical analysis. In addition, the experimental

---

*Paper presented at the Seventh Annual Phi Delta Kappa Symposium on Educational Research, Madison, Wisconsin, August 11, 1965.

material itself was highly susceptible to what we now call treatments, in that one can apply a given number of pounds of potash per acre, feed pigs a given ratio of corn to milk, or grow plants of defined genetic structure. Thus, a situation existed in which one had experimental material which could be measured and at the same time susceptible to manipulation by means of outside agents. Such was the situation less than half a century ago, when Sir R. A. Fisher began the development of modern statistics and experimental design. Fisher was chief statistician at the Rothamsted Experimental Station, and it was in this context that he developed the basic concepts of experimental design which were later imbedded in his classic book, "Design of Experiments." In this remarkable book he presented the logical basis for statistical inference, the fundamental concepts of randomization, replication, and local control, and various techniques for laying out or designing experiments, as well as a number of related statistical techniques, thus essentially defining the modern field of experimental design. During the past thirty years, both theoretical and applied statisticians as well as research workers in diverse disciplines, have extended, refined, and publicized the original contributions made by Fisher. Even the most recent texts on experimental design are devoted almost entirely to topics which stem directly from Fisher's work.

One of the fundamental principles underlying Fisher's experimental designs was that they should be self contained in order that a valid estimator of the error variance be available for use in testing hypotheses and/or purposes of estimation. The application of this principle in conjunction with the basic characteristics of agricultural research of that period resulted in experimental designs for what I shall call "stand alone" experiments. By this I mean that each experiment was designed as a logical entity without particular concern for any other experiment.

For example, the experimental design of a corn growing experiment would not directly be influenced by a concurrent experiment on hog breeding.

That is not to say that the experiments themselves were not related, but rather that the statistician did not take the other experiment into direct account when designing the particular research at hand. These "stand alone" experiments were typically conducted by a single investigator aided by a number of assistants who would tend the land, apply the treatments, harvest the crop, and collect the necessary measurements. The principal investigator would perform the analysis and write a paper to report his results to the scientific community. Upon completion of the experiment the investigator would proceed to another stand alone experiment which might or might not be related to the experiment just completed. Recent developments by Box (1954) and others have extended this experimental design approach so that certain types of experimentation are not necessarily stand alone in character, but for the moment we shall by-pass the recent developments and defer them to a later point in the present paper. The "stand alone" experiments and the experimental designs which Fisher developed for them have a strong similarity to laboratory experiments in which one has a reasonably well-defined problem, experimental materials which can be measured, and variables which are amenable to manipulation. If one peruses the published research in disciplines such as psychology and education, it can easily be seen that "stand alone" experiments designed in accordance with Fisher's principles account for a large proportion of current research. The degree to which Fisher's work has permeated the scientific community can be ascribed in part to the characteristic of the research workers who were sophisticated enough to see the significance of Fisher's contributions to disciplines other than agriculture. Such persons were typically, though not necessarily restricted to, university professors

conducting individual research on extremely limited funds. Hence, a system of experimental design applicable to small scale research which would maximize scientific information per unit of cost was extremely welcome. One should also note that during a good portion of the period under discussion the world was either in the depths of a severe economic depression or in the throes of a major war, hence monies for basic research were dear. Even such a distinguished researcher as Professor Palmer Johnson, who founded these Phi Delta Kappa Research Symposia considered himself fortunate to obtain a modest grant providing a single half-time graduate assistant and five hundred dollars with which to conduct an experimental study. It should be noted that many of Fisher's own later developments as well as those of others were motivated by the desire to maximize the return at a minimum cost. Incomplete factorial designs, for example, essentially yield particular results of a complex design within the cost limitations of simpler designs at the expense of certain types of information. Thus, even the economics of the time can to some degree be considered to have contributed to the popularity of the "stand alone" experiment.

To the present author it appears that experimental design as we now know it grew out of the situation in which an individual research worker performed relatively small scale experiments which were logical entities. It should not be implied, however, that the experiments were not sophisticated, as there is no necessary connection between the size of an experiment and its level of sophistication. If any relation does exist I suspect that it is an inverse one. The individual research worker and the stand alone experiment designed in accordance with Fisher's principles will continue to be the backbone of scientific progress, yet recent events have conspired to change the fundamental character of a large segment of present

and future educational research activity. The cause of this change is the increasing availability of funds for educational research, primarily from the federal government.

The availability of significant amounts of federal money for research is a phenomena which began in the physical sciences during and after World War II. Support has come from agencies such as the National Science Foundation, the Office of Naval Research, and the Atomic Energy Commission. The appearance of federal monies for educational research is of relatively recent vintage beginning with the establishment of the Cooperative Research Branch of the U. S. Office of Education in 1953 and not really getting started until 1957. In the early days of this program, only a modest amount of money was available for educational research but through considerable effort on the part of many persons this program is currently funded at about 25 million dollars per year. Due to the passage of recent legislation additional monies in considerable amounts are availabe from a number of different programs within the Department of Health, Education and Welfare.

In recent years the total amount of monies available has increased rather suddenly without any corresponding increase in the number of educational researchers competent to spend these monies, thus, creating a serious problem. Anyone who has dealt with bureaucracy knows that unspent appropriations are a sure sign of incompetence and that the solution is a rapid application of Parkinson's Laws. Thus, the fastest way to allocate money is to fund larger, more expensive projects. Despite the obvious Parkinson's Law aspects, the emergence of the large-scale research project is primarily due to the belief that today's complex educational-social

problems cannot be effectively studied in a piecemeal fashion by individual investigators. The opinion held is that such complex problems are best studied by a team of researchers, developers, and disseminators in possession of sufficient funds and facilities over a reasonable period of time. For purposes of the present paper I shall beg the question whether this is the best or even the only approach, but will assume that it is a current approach.

Accepting the premise that in the foreseeable future the trend is towards heavily funded research involving a large-scale project, let us indicate what the characteristics of such research are likely to be. The research problem itself will be on some broad problem area such as cultural deprivation, teacher effectiveness, etc., rather than on some specific problem such as reading rate or serial learning. Because of the breadth of the problem an interdisciplinary approach involving disciplines such as psychology, sociology, political science, economics, and medicine, as well as education, will be used to attack the problem area from a number of points of view. An additional facet of large scale research projects is that because of the sheer number of persons involved problems of management inevitably arise.

From the above it can be seen that the research milieu in which present educational research is being conducted is quite different from that for which Fisher developed classical experimental design. Although Fisherian experimental designs are certainly still valid, the rate of change in classical experimental design has not kept pace with the rate of change in the areas for which the educational researcher needs experimental designs. Undoubtedly the milieu in

which the agricultural researcher of the 1920's operated appeared to him as complex, confused, and intractable as does the current situation, yet out of his era classical experimental design evolved. The thesis of the present paper is that the current situation presents an opportunity for a new conceptualization of experimental design which is equally as great as that which existed during the 1920's.

Some Characteristics of Large Scale Research Projects

In order to ascertain what the specifications for this new class of experimental designs might be, the characteristics of large-scale research projects germane to experimental design considerations are discussed in the paragraphs below.

Strange as it may seem, one of the most difficult tasks associated with large scale research projects is that of defining the problem. There appears to be an inverse relationship between the dollar amount of a grant and the amount of specificity in the statement of the problem. A small scale project is far more likely to be concerned with a problem which can be investigated via a "stand alone" experiment, whereas the large-scale project is more likely to be defined in terms of some problem of broad educational or social concern. For example, an area receiving much attention is that of cultural deprivation. Because cultural deprivation is a compound of social, educational, and economic factors, it is difficult to specify problems in this area in other than broad terms. Many other areas in education such as teacher effectiveness and creativity are also complexes of many factors, and understanding any one of them does not unravel the total

problem area. Admittedly, in any of these complex fields it is possible to define specific problems which are amenable to classical experimentation, however, such specific projects would not normally be funded at a level which would label them as large scale research projects. The contrast in the level of specificity in the current situation and that of the agricultural context is striking. In the latter the problem was very specific; is a differential effect in the number of bushels per acre due to several levels of fertilizer concentration? In the former, on the other hand, it is very broad; for example, can one provide an environment which enables children to overcome the effects of cultural deprivation? How to design an experiment for the specific is straightforward; how to design an experiment for the ill-defined is not. Part of the difficulty in problem definition stems from a somewhat unclearly drawn boundary between scientific research and implementation of educational change. The educational researcher who considers himself a basic research worker is interested in studying problems with a view towards understanding the principles, the processes, and the dynamics of a problem area. The understandings gained eventually result in a better theoretical framework for the problem area, from which one secures better conceptualizations for research in that area. The educational researcher who implements educational change is concerned with what can be done to alleviate a social ill or correct a deficiency in the education system as rapidly as possible. To him the problem, though not defined with great precision, is an obvious one and readily identifiable. If a particular programmatic change seems to produce desirable results, he is satisfied that the goals of his research have been obtained. These two points of view are not dichotomous because they lie on a continuum,

but it is because of this continuum that the design of experiments is so difficult.

The problem of specificity appears in another comparison of current and past research milieus. In the agricultural experiment the dependent variable or criterion measured was specific, such as the number of bushels per acre or the animal weight at slaughter. The existence of a specific measure makes mathematical analysis possible and because of Fisher's genius also quite easy. In the large-scale research project an obviously relevant criterion variable may or may not exist and if it does, its specificity leaves much to be desired. Because the problem is broadly defined it is not usually possible to reduce the criterion to some specific measure such as reading score or GPA. Rather, one is concerned with increased educational potential, improved social adequacy, or any of a number of criteria which are equally difficult to define and exceedingly difficult to quantify. One, of course, can always devise instruments which purport to measure such global variables, but the development of these instruments is a larger project than the research projects under discussion. There are of course some existing possibilities in constructing linear composites of variables and using the composite, but none the less a serious problem exists in large scale research projects in regard to defining and measuring adequate criterion variables.

When an interdisciplinary approach is taken, simultaneous research occurs in several disciplines within the same research framework, and one is faced with the problem of how to utilize the measures obtained in the several areas to analyze the problem as a whole. Multivariate analysis is possible if all the measures are collected within a given experiment, but that is not the situation envisioned. The measures would have been taken in different experiments

at different times for different reasons, yet they all contribute to understanding the problem area as a whole. The interdisciplinary approach of large scale research problems is an important characteristic and adequate means for effectively exploiting it must be found.

Thus, the large-scale research project has unique characteristics such as, an ill-defined problem, the lack of an obvious criterion variable, and the necessity to integrate results from several disciplines, which differentiate it from the classical laboratory type experiment.

Application of Current Design Procedures to Large Scale Research Projects

Within the framework of existing experimental design procedures a number of approaches to designing large scale research projects are possible and several are discussed below.

First is the employment of one of the classical designs, and the project is then devoted to performing the operations required by the treatment level combinations which must appear in the cells of the design. Within this classical approach, one has two paths which can be taken. (a) Make the design extremely large, complex, and involved, then the energy of the project is devoted to trying to accomplish this complex design. Unfortunately, in the educational context, complex designs with many treatment levels etc., become extremely difficult to manage and can easily degenerate. (b) Use a simple classical design in the programmatic variation situation where the treatment levels are the various programmatic changes such as curricular innovations. The difficulty, of course, is that the treatment levels are specified in terms which are as broad as those used to describe the problem. The classical experimental design in this situation permits one to test hypotheses about the programmatic variation but it does not permit one to ascertain analytically the actual bases for the obtained results.

A second approach to experimental designs for large-scale research projects is a function of the interdisciplinary attack which is currently held in high esteem. The total problem can be fractionated into numerous smaller experiments in the several disciplines involved. These smaller experiments can then be designed in accordance with classical principles and conducted as "stand alone" experiments which are in some sense related to the total project. Within the interdisciplinary context this approach is attractive, as it permits each investigator to perform research within the area of his own competency, using his familiar tools, and he need not be overly concerned with interaction of his experiments with those of others. Although this approach provides the individual investigator within the project a great deal of freedom, it does make it difficult to perceive the accomplishments of the project. The integration of the results of these numerous stand alone experiments must be accomplished by the project director employing his clinical judgment as to the contribution of each piece of research to the goals of the project. Integration of this clinical type is not simple even in small scale projects and appears to be nearly impossible in the case of large projects dealing with rather global types of problems such as described above.

The third and least desirable approach is no design at all. Because of the global nature of the problem, the diversity of disciplines from which the problem can be viewed, and the lack of suitable guidelines, large-scale research projects can proceed without any explicit experimental design at all. Such projects are much like the military study contracts in which the vendor is paid to probe in a particular area and see if he can find anything interesting. Educational research projects can also be of this "let's see if we can find anything" variety. The curriculum is varied, special classes for parents are established, television is employed, etc., and one "observes the effects."

While occasionally garnering considerable publicity and revealing a great deal of innovative ingenuity, the contributions of such projects are limited. The lack of adequate criteria and the inability to test hypotheses under these conditions severely limit the usefulness of such projects.

Although specific examples of these three approaches to laying out large scale projects have not been given, illustrative examples could easily be drawn from the lists of projects funded by any of several agencies. Each of the above approaches has its advantages and disadvantages, but in the view of the present author none of them appear to be adequate to meet the design needs of a large-scale project dealing with a complex problem.

In the paragraphs above characteristics of large scale research projects which need to be considered when creating a new class of experimental designs have been presented. In order to clarify these points and also impart somewhat the flavor of how in the absence of adequate experimental designs an investigator proceeds in this context, one of the author's own projects will serve as an example. Although this is not a large scale project it does possess a number of the necessary characteristics.

The Concept Attainment Simulation Experiment (CASE) is a project in which the technology of computer programming is being used to study concept learning by means of simulation programs. The ultimate goal of this project is that the computer program exhibit concept learning behavior analogous to that of human subjects learning the same concepts. Such correspondence is not simple due to the variability of a given subject when learning a concept and the considerable variation between subjects. For purposes of the project any reasonable approximation to human behavior in this problem solving context would be acceptable. Thus, although the goal of the project is quite clear, how to attain that goal is not clear at all and because of this, it was not

possible to design a classical experiment to reach its goals. Instead a heuristic approach was taken which involved two techniques. First a long-range plan was drawn up on essentially a priori grounds as to the kinds of areas in which one would need to investigate and a reasonable sequence in which the areas would be investigated. The long-range plan specified at what point in the project each area would need to be studied and why these studies would contribute to the total project. The subsidiary studies conducted as part of the long-range plan have a distinct building-block nature as adequate conceptualization of studies to be performed six months from now depend upon the results of those currently being conducted. In many cases these auxiliary projects are in fact "stand alone" experiments designed to test particular hypotheses which the long-range plan indicates will need to be tested before proceeding. The existence of a long-range plan serves as a mechanism by which one can anticipate future needs and try to produce information which will meet these future needs.

The second heuristic technique was a "target of opportunity" approach in which promising leads are followed when they appeared, even if they were not planned. For example, very early in the CASE project it became quite clear that subjects differed in the basis upon which they selected objects from their external environment. A series of short studies revealed a dominant attribute phenomenon in which some subjects were attracted by color, others by shape, etc. The simulation program was subsequently modified to incorporate a mechanism by which it exhibited its own dominant attribute behavior. The provision for the "target of opportunity" type study protects the project from being unreasonably confined by the long-range project plans. Conversely, the long-range project plan provides protection against the project's degenerating into a series of sub-projects on inviting but unproductive avenues of research.

These heuristic techniques have enabled the concept simulation project to proceed at a reasonable pace, but there is not any coherent design to the total experiment by which one can determine if any particular facet contributes to the whole or if the whole is in fact contributing to our understanding of concept learning processes. What appears to be required is an experimental design appropriate to such projects in which scientific rigor would replace the heuristics presently employed and one would proceed on grounds other than clinical judgment. It is not clear to the present author how any of what we currently call experimental design could be directly or indirectly applied to a project such as that described above.

Section III

Characteristics to be Possessed by the New Experimental Designs

An important distinction one might make between milieu of the present and that of the 1920's is that the current interest is in investigating a problem area rather than a specific problem within an area. Whereas classical experimental design provided a paradigm for examining a particular problem, the new design must provide one for investigating a large complex area of interest. The paragraphs below describe the characteristics which the present author feels the new experimental designs should possess in order that one can implement for large scale research projects dealing with problem areas.

A significant characteristic of classical experimental design is that it dictates certain principles of experimentation such as randomization, replication, and local control which must be adhered to if valid results are to be obtained from the experiment. These principles coupled with the layout or the design of the experiment provide a framework within which the investigator conducts his experiment. Similarly, the new experimental design must

provide principles of experimentation and layout which serve as a framework for large-scale research projects. Unless the experimental-design system provides such a framework, the research-management hierarchy will be unable to adequately perform their decision-making function within a project.

The new experimental design must be inherently dynamic and possess the ability to change its internal structure without sacrificing the rigor of the design. The simulation research project described above was characterized by its fluid internal state in which ideas, insights, and understanding could be generated which would significantly alter the research during the course of the project. Doyle (1965) has indicated, "as we come closer to the basic research end of the spectrum, however, it becomes more and more imperative to be free to alter the plan. Indeed, in basic research altering the plan ought to be a state of mind."

The current interest in the interdisciplinary approach indicates that the new experimental design must possess a capability for conducting "stand alone" experiments in the several disciplines and yet integrate the results of these experiments into the main plan of the experiment. An analogy exists in the systems approach used in the military to develop complex weapon systems. Separate companies develop a navigation system, the weapon, the fire control system, and the delivery vehicle. Although each area involved different skills technology, and variables, the final products are integrated into a functioning weapon system. Similarly the experimental design should permit research to be conducted in the several disciplines associated with a problem area, yet insure that each subsidiary experiment can make its unique contribution to the total project.

The capability for allowing subsidiary experiments to be a facet of the total design, coupled with the capability to alter the plan, seems to imply an

unusual hypothesis-testing scheme. Within the subsidiary stand-alone experiments the classical procedures are used to test specific hypotheses. However, in the main body of the design, hypotheses must be tested to determine whether the results of the stand-alone projects are to affect the total project. The main experimental design is a common thread to which the stand-alone experiments are attached. The results of these experiments can alter this common thread in various ways or merely indicate continuation along the original path. The peculiarity in hypothesis testing arises in that hypotheses tested relevant to the common thread or plan of the experiment may not be based upon specific data or variables collected for that hypothesis, but could be based upon the pattern of results of hypothesis tests performed within the auxiliary "stand alone" experiments conducted in the several disciplines. One could envision a hierarchical structure of hypothesis tests in which one moves about a complex decision tree containing the possible paths where results of the specific sub-hypotheses will determine the subsequent branches to be followed. Returning to the cultural deprivation example again: separate experiments could be conducted on, say, reading rates of children and adaptability of children to social stress. In each case hypotheses about particular treatments using particular variables can be tested. The results of these experiments contribute to testing hypotheses about some higher-level concept of importance to cultural deprivation. If this higher-level hypothesis is rejected, one avenue of research is pursued and if not rejected, another is followed. Thus, the hypothesis-testing procedure reflects and implements the dynamic nature of the new experimental design required by large-scale research projects.

Implicit in all of the above characteristics is that the experimental design provide the overall plan for the life of the experiment.

Without experimental design the project managers cannot make the day-to-day decisions necessary to keep a project on an even keel, the research worker in a specific academic discipline cannot perceive how his work contributes to the total project, and the scientific rigor of a complex project cannot be maintained. Thus, the new experimental design must provide the necessary structure within which one conducts research, just as presently provided by classical experimental design.

The characteristics of experimental design required by large-scale research projects may seem curiously confounded with what we currently consider to be research planning and management, and this is certainly true. It is just this truth which is the emphasis of the present paper, namely, that one cannot conduct large-scale research without an experimental design which provides scientific rigor to these aspects of research.

Section IV

Existing Bases for the New Class of Experimental Designs

The immediate creation of a class of experimental designs to meet the needs of large-scale research projects is of course not possible. One must attempt, however, to find useful concepts and approaches to existing techniques which can be employed in the development of the new experimental designs. There are several existing areas of investigation which appear to offer a basis for the necessary developments. These are the Program Evaluation and Review Technique (PERT) of the U. S. Navy's Polaris Project, Professor Box's response surface designs, and the general field known as operations research (OR). Each of these fields has unique characteristics which can be

exploited. The paragraphs below indicate the present author's notions as to their possible contributions.

PERT

The Program Evaluation and Review Technique is a management tool used to keep large-scale military development programs proceeding according to schedule. The large scale project is fractionated into its component activities and a network developed which depicts the relationships existing among the multitudinous activities. The PERT network normally employs only a single variable, which can be either time (PERT/time) or cost (PERT/cost). To use time as the case in point, the date at which the total project must be completed is ascertained and the amount of time necessary to accomplish each of the activities specified in the network is also ascertained. The path through this network requiring the maximum amount of time is the critical path, and if the total time along this critical path exceeds the completion date the project goals will not be met. In order to meet the completion date, the times allocated to the activities within the network must be reassigned or the network itself can be redesigned. A computer program is employed to analyze the network, and the reallocation process can be repeated until a satisfactory critical path is obtained. The PERT network is a very dynamic entity due to the various activities requiring times other than that allocated to them and as each activity is completed, or not completed, the network changes. Although PERT should play a major role in the management of a very large-scale research project, the present author is not interested in the management aspect. The unique feature of PERT of interest is its dynamic network which permits the internal structure of the project to vary over wide bounds, yet the goals of the total project are satisfied. Admittedly, a

single variable such as time or cost is the criterion measure, but the dynamic-network concept could be extended to provide a mechanism by which the overall plan or "common thread" of a project could be maintained. Perhaps the hierarchical structure of hypothesis testing could be incorporated into a PERT type network analysis. The second facet of the PERT approach is that it permits extremely divergent activities to proceed in a series, parallel, and a series parallel fashion, yet all of these activities are integrated into a meaningful whole, namely, the project. These two features of the PERT approach, dynamic network analysis and integration of diverse activities into a meaningful project, appear to offer much upon which the new experimental design can be built.

Response Surface Designs

The response surface designs due to Box (1954) are an ingenious extension of the classical stand-alone experimental design which overcome, in a certain sense, limitations mentioned earlier that the results of one experiment do not influence the design of the next. In the response surface situation one has some response say, yield, which is connected to a group of k quantitative variables or factors such as temperature, concentration, and time. One is interested in finding the set of conditions i.e., levels of the factors, which will optimize the yield or response variable. The relationship existing between the yield and the several factors can be plotted as a surface in an appropriately dimensioned space. The essential problem is to allocate the cells of a classical experimental design in this space in such a way as to ascertain the location of the maximum point, plane, or ridge of the response surface. Usually the original points do not produce the maximum but they do provide clues as to the possible location of the maximum. Based upon these

clues, new levels of the factors are established and additional cells of the experiment are performed to locate the maximum of the response surface. The analysis is based upon a multiple regression approach in which a surface of specified dimension is fitted to the experimental points. The advantage of the response surface design is that it is exploratory in nature. It assumes that the experimenter knows enough about his area to roughly locate the factor levels near the optimum yield, then response-surface design provides a mechanism by which experiments can be conducted to direct the experimenter to the factor levels which yield a maximum.

Response surface designs have the same "laboratory" quality about them that are possessed by the classical Fisherian experimental designs, namely, that they presuppose interval-scale factors such as temperature, pressure, etc. which are amenable to manipulation and an easily measured response variable. However, the exploratory capability of these designs is a characteristic which should be of great value to new designs for large-scale research projects. The exploratory characteristics would need to be expanded, along the lines indicated by Box (1954) to designs involving multiple response variables such as required by the interdisciplinary approach and to factors which are of interest to social scientists rather than the physical scientist.

Operations Research

The third and final area of interest is that of operations research, a field which grew out of the early attempts of the British to effectively use radar, which they had invented just prior to World War II. The essential characteristics of operations research as given by Askoff and Rivett (1963) are (1) the systems orientation, (2) the use of interdisciplinary teams, (3) and the adaptation of the scientific method. The early experience involving

military problems associated with radar showed that the actual and stated problem rarely coincided, hence it was necessary to investigate beyond the scope of the stated problem area. In order to do this the operations research people found it necessary to study and analyze the total context within which the stated problem occurred. The study of this total context and of the relationships of the important variables within this context has become known as the systems approach. Because of the manpower shortages during World War II, the early operations research groups acquired persons with a wide diversity of training ranging from mathematicians to psychologists and even some medical specialists thus inadvertently forming an interdisciplinary group. The interdisciplinary approach proved so successful in providing new ways of looking at problems that it has become an integral part of the operations-research approach. It is evidenced primarily by the fact that variables from both the physical and social sciences are included in analyses when appropriate.

Askoff and Rivett (1963) stated that the basic equation of all operation research models is $P = f(C_i, U_j)$ where P is performance, C_i are the controlled variables, and U_j are the uncontrolled variables. The C_i are those variables whose levels can be manipulated much in the same fashion as the treatments or factors in classical experimental designs. The U_j are factors of which one is cognizant but yet are beyond direct manipulation, analogous to the variables in education which get lumped into error variance. The development of an adequate measure of performance is the most difficult aspect of this relationship, and in industry it usually becomes a function of cost. The goal of the operations research approach is to optimize (either maximize or minimize) performance, a goal strikingly similar to that of response-surface designs. The third aspect of operations research is that it is concerned with developing models. The real-world problem areas within which the operations

research specialist operates are not ordinarily amenable to direct manipulation, for the changing of certain variables could lead to financial ruin if the investigator happens to be wrong. Hence, operation researchers attempt to develop mathematical models, simulation models, and other types of models which permit them to manipulate the model and ascertain outcomes without great danger. Of course the eventual results must be applied in the real world, but model building and manipulation reduce the attendant risks.

From a mathematical point of view, the fundamental equation of operations research is expressed as a system of linear relationships involving inequalities rather than equalities and an objective function expressed in units of say, cost, which is to be optimized. Thus, a system of n "equations" are to be solved, n minus one of these involving inequalities and the objective function involving an equality. It should be noted that all terms in these equations are of the first order, in contrast to the use of higher order terms in a response surface design. The industrial applications of operations research technology is extensive, and problems of a wide variety of forms and content are studied via this technique. A formidable literature that is far beyond the scope of the present paper to review exists in this area; rather, the concern is for the features of this technology which can be adapted to experimental design. From a positive point of view, operations research workers appear to have had a great deal of experience in the utilization of interdisciplinary teams of researchers, and this experience should be exploited in developing experimental designs involving various disciplines. The general systems approach associated with operations research appears to the present author to be a critical ingredient of any large-scale research effort.

It should also be noted that the PERT approach described above is merely a special case of the operations research approach and is one of the

outstanding achievements of this field. The idea of expressing constraints placed upon the relations of the variables as inequalities does not seem to appear in the classical experimental literature and perhaps should be a consideration. Finally, the concept of developing models for complex systems prior to actually manipulating them, is something which all areas aspire to, but in educational research, at least, few if any models exist which approach the level of sophistication used in the operations-research area.

Operations research offers much in the way of approaches to problem areas via the systems orientation and the specialized mathematical methods for solving large systems of equations via numerical techniques. Despite these features, it does not provide experimental designs for research in the areas to which operations research is applied.

Summary and Conclusions

Classical experimental design as originated by Fisher grew out of the agricultural research milieu of the 1920's and 30's and was developed to meet the design needs of research workers of this era. It has been thirty years since Fisher's Design of Experiments first appeared, and it was the basic premise of the present paper that today's research milieu, especially in education, is considerably different and hence offers the possibility for developing a new class of experiment designs as uniquely suited to today's problems as classical design was for the problems of the 1920's.

The particular feature of today's situation that clearly distinguishes it from the earlier period is the existence of the large-scale research projects which have resulted from the availability of significant amounts of federal money for educational research. Today's large-scale research projects possess a number of characteristics: the project deals with a problem area

rather than a specific problem, the problem itself is ill-defined, a univariate criterion variable such as yield or cost is not readily available, an interdisciplinary team of researchers are involved, and a management hierarchy exists which needs research guidelines; all of which differentiates it from those of an earlier era. Classical experimental designs do not appear to possess the capabilities necessary to cope with research projects possessing these characteristics.

From the point of view of the present author a new class of experimental designs which possess the following characteristics needs to be developed:

(a) The experimental design should provide the framework or common thread which serves the guideline for the conduct of the experiment over its life span. It should also provide principles of experimentation upon which the research administrator can base decisions.

(b) The experimental design should be dynamic in nature in that it can allow for alterations in the design as a result of information acquired during the course of the design. In other words, the design is not a fixed plan for the experiment but rather is a strategy for conducting a large-scale research project.

(c) The interdisciplinary approach suggests that the experimental design possess an integrative function which will permit subsidiary experiments to be conducted in various disciplines yet contribute to the total design. A function of this integrative function appears to be some form of network analysis involving hypothesis testing as a decision-making device within the design. One should be able to fractionate a problem area into related sub-systems and have the experimental design provide a scientifically rigorous scheme for integrating the end products of the sub-systems into the framework of the total project.

The present author does not have a clearly conceived idea as to how one would create a class of experimental designs possessing the above characteristics but has only tried to indicate the type of challenge which is presented by today's research milieu. Because of the lack of a clearly perceived future, features of existing techniques such as Operations Research, PERT, and response surface designs were examined for useful characteristics. The general systems orientation and interdisciplinary approach of Operations Research contains much of what is desired; however, the most attractive features of Operations Research are also those which are the least formally developed. The response-surface designs, especially in their multivariate extension, most closely approach what is desired and appear to offer the greatest possibility for immediate developments leading to the new experimental design.

In order to assure the reader that the present paper is not advocating large-scale research projects, I would like to close with a short quote from an editorial by Vannevar Bush (1963) which puts many large-scale projects in proper prospective. He states: "The spectacular success of applied research during the war led to a fallacy entertained by many. It is that any problem can be solved by gathering enough scientists and giving them enough money. To solve the problem of the common cold assemble a great institution, fill it with scientists and money, and soon we will have no more colds! The great scientific steps forward originates in the minds of gifted scientists, not in the mind of promoters. The best way to proceed is to be sure that really inspired scientists have what they need to work with, and leave them alone."

References

Ackoff, R. L. and Rivett, P. A. A Manager's Guide to Operations Research. New York: John Wiley and Sons, 1963, 107 pp.

Box, G. E. P. "The Exploration and Exploitation of Response Surfaces: Some General Considerations and Examples." Biometrics, Vol. 1954, 16-60.

Bush, U. Editorial, Science, 142, 1963, 1623.

Doyle, L. B. "Seven Ways to Inhibit Creative Research" Datamation, February 1955.

JUL 1 2 1967

Reprinted from

EDUCATIONAL and PSYCHOLOGICAL MEASUREMENT

A QUARTERLY JOURNAL DEVOTED TO THE DEVELOPMENT AND APPLICATION OF MEASURES OF INDIVIDUAL DIFFERENCES

Correlation as a Function of Predictor Score Points. NAMBURY S. RAJU AND ISAIAH GUTTMAN 655

Predicting Achievement in Ninth-Grade Mathematics from Measures of Intellectual-Aptitude Factors. J. P. GUILFORD, RALPH HOEPFNER, AND HUGH PETERSON .. 659

Difficulty and Other Correlates of Criticalness Response Style at the Item Level. LAWRENCE J. STRICKER 683

An Analysis of Test-Wiseness. JASON MILLMAN, CAROL H. BISHOP, AND ROBERT EBEL 707

The Effects of Guessing on the Standard Error of Measurement and the Reliability of Test Scores. DALE MATTSON 727

The Interaction of Personality and Intelligence in Task Performance. DAVID KIPNIS AND CARL WAGNER 731

The Social Desirability Variable: Implications for Test Reliability and Validity. ALFRED B. HEILBRUN, JR. 745

(Continued on inside front cover)

VOLUME TWENTY-FIVE, NUMBER THREE, AUTUMN, 1965